

Low-cost and Faster Tracking Systems Using Core-sets for Pose-Estimation

Soliman Nasser, Ibrahim Jubran, and Dan Feldman

Robotics & Big Data Lab,
Computer Science Department,
University of Haifa, Israel

Abstract. In the pose-estimation problem we need to align a set of n markers (points in 3D space) and choose one of their $n!$ permutations, so that the sum of squared corresponding distances to another ordered set of n markers is minimized. We prove that *every* set has a weighted subset (core-set) of constant size (independent of n), such that computing the optimal orientation of the small core-set would yield *exactly* the same result as using the full set of n markers. A deterministic algorithm for computing this core-set in $O(n)$ time is provided, using the Caratheodory Theorem from computational geometry. We can then boost the performance of inefficient algorithms or popular heuristics (such as ICP) by running them (maybe many times) on our small coresets.

Based on this coresets, we developed a low-cost ($< \$100$) real-time tracking system. As an example application, we turn a toy quadcopter ($< \$20$) with no sensors into an “autonomous drone” that is controlled by a mini-computer ($< \$30$). Experimental results show that, unlike uniform sampling of features, our suggested coresets yields error that is comparable to commercial ($> \$10,000$) tracking systems for such simple but common applications. Open source code is provided [1].

1 Motivation

How can we compute the orientation of a robot or a rigid body in space? This is a fundamental question in SLAM (Simultaneous Localization And Mapping) and computer vision [21, 13, 18]. For the case of static ground cameras outside the robot, the standard approach is to put a few markers on the robot or object in known places and distances between them, and then track them in real-time using few cameras by triangulation [14]. Instead of markers in $d = 3$ dimensional space, visual features, e.g SIFT or SURF [19, 4] in high d -dimensional space can be used.

The same problem arises when the cameras are on an autonomous moving robot and few points (markers/features) are observed. If the position of these points in the real world is known from a map or a 3D-model, the robot can compute its position by matching them with the observed set of points.

By letting $P = \{p_1, \dots, p_n\}$ denote the initial/known set of $|P| = n$ points in \mathbb{R}^d , and letting $Q \subseteq \mathbb{R}^d$ denote the observed points (after translation and rotation of the object), the localization problem can be divided into three sub-problems that in some sense are independent:

Detection of the set of n markers $Q = \{q_1, \dots, q_n\}$ in the observed image.

Matching which point $q = f(p)$ in the observed set Q corresponds to each point p in the initial set P . In general, there are such $n!$ such permutations (matchings) $f : P \rightarrow Q$, where $f(p) = q$.

Pose Estimation of the set Q , i.e., compute the rotation matrix $R \in \mathbb{R}^{d \times d}$ and translation $\mu \in \mathbb{R}^d$ of P that yields the observed set Q . The matrix R and the vector μ then corresponds to the position and orientation of the robot, compared to its original position. In practice, there is no exact matching due to noise in the observed set such as latency of camera, resolution, frames per seconds, blurring, communication speed and quality, etc. Thus Gaussian noise is usually assumed, which implies that the sum of squared distances (RMS error) between each point in $p_i \in P$ to its corresponding shifted point $f(p_i) = q_i$ in Q should be minimized,

$$\text{OPT}(P, Q) := \min_{R, \mu} \sum_{i=1}^n \|p_i - (\mu + R(q_i))\|^2; \quad (1)$$

see Definition 1.

2 Related Work

The pose estimation problem is also called the alignment problem, since given two 3D point sets, P and Q , the task is to find the Euclidean motion that brings P into the best possible alignment with Q . The optimal translation μ^* is simply the mean of Q minus the means of P , each can be computed in $O(nd)$ time. Computing the optimal rotation R^* requires computing the d -by- d matrix $P^T Q$ in $O(nd^2)$ time where each marker is a row in P or Q , and then computing the Singular Value Decomposition (SVD) of this matrix in $O(d^3)$ time [12].

The standard and popular solution for solving the matching and pose-estimation problems is called Iterative Closest Point (ICP) proposed by Besl and McKay [5]; see [24] and references therein. This algorithm starts with a random matching (mapping) between P and Q , then (1) runs the Kabsch algorithm on this pair, (2) re-matches each marker in P to its nearest point in Q , then returns to step (1). Indeed, we run these steps on our coreset below which allows us faster tracking using low-cost hardware. Various data structures, like k-D tree [11] or spatial bins [25], are used to facilitate search of the closest point. To speed up the convergence, normal vectors are considered, which is mainly helpful in the beginning of the iteration process [20].

References to more provable approaches that try to compute the minimum Hausdorff distance under Euclidean motion from can be found e.g. in [7]. These techniques from computational geometry are relatively slow, roughly $O(n^5)$. Running such algorithms on our suggested coreset below might turn them into practical heuristics that are based on provable solutions.

3 Novel Approach: Core-set for Pose Estimation

In order to boost the performance of each of the three steps above (detection, matching and pose-estimation), our main theoretical result is an algorithm that computes in $O(n)$

time a “smart selection” (core-set) of a small subset \tilde{P} of the n markers in P and their corresponding markers \tilde{Q} in Q , with appropriate multiplicative weights. Maybe surprisingly, we prove that for *every* such a pair (P, Q) of sets we can compute a pair (\tilde{P}, \tilde{Q}) of core-sets that have the following guarantees (see Theorem 6 for details):

- (i) The size $|\tilde{P}| = |\tilde{Q}| = O(1)$ of each coreset is constant, i.e., independent of n .
- (ii) The rotation R and translation μ that minimize the sum of squared distances from \tilde{P} to \tilde{Q} , would also *exactly* minimize this sum between the original sets. In particular, $\text{OPT}(P, Q) = \text{OPT}(\tilde{P}, \tilde{Q})$.
- (iii) The same coreset and the above guarantees remain the same and thus can be used also after the observed set Q is rotated by R or shifted by μ ,

$$\text{OPT}(\tilde{P}, R(\tilde{Q}) + \mu) = \text{OPT}(P, R(Q) + \mu).$$

- (iv) The coresets (\tilde{P}, \tilde{Q}) can be computed in $O(n)$ time using one pass over the input points and $O(1)$ -memory.

The constant behind the $O(1)$ notation is bounded by $r(d-1) + 1 = O(d^2)$ where r is the rank of P and d is its dimension.

For example, in our experiments where n IR markers are put on a quadcopter such that $r = 2$ as in Fig. 2(left)), the coreset is of size $r(d-1) + 1 = 2 \cdot 2 + 1 = 5$. This enabled us to obtain significant improvements even for very small values $n \sim 10$, partially from the reasons that are explained in the next section.

Unlike uniform random sample of markers, that are used e.g. in RANSAC-type algorithms [8], our core-set is *guaranteed* to produce an optimal rotation matrix R^* for the *original pair* (P, Q) using only the small coreset pair (\tilde{P}, \tilde{Q}) . This theoretical guarantee turns into practical results in our experiments and videos when the error scale is changed from meters to centimeters.

4 Why core-set for Pose Estimation?

Pose Estimation. Given the above coresets (\tilde{P}, \tilde{Q}) , we can solve the pose-estimation problem in $O(1)$ time by solving it on the coreset. Property (iv) allows us to compute the coreset in $O(n)$ time, which is the same time that is needed to solve the pose-estimation problem. However, in the next observed frames we can use Property (iii) of the coresets above, and reuse them also for the following frames, while the the set Q of markers on the observed rigid body keeps moving. This reduces the computation time for the pose-estimation problem in each frame from $O(n)$ to $O(1)$.

While our coreset is designed for the pose-estimation problem, the fact that the number of tracked markers reduces from n to $O(1)$, allows us to handle efficiently also the solution to the following pre-processing problems for the pose estimation that are mentioned in Section 1:

Detection of the set of n markers in the observed image is reduced to detecting of only the required markers in the coreset.

Matching the n points in P to the n points in Q from the $n!$ permutations now reduces to choosing one among constant number of permutations.

The improvement in the running time and performance of these first two steps prior to solving the pose-estimation problem, depends on their exact implementation. For example, if we apply the ICP heuristic for solving this problem on the coreset instead of the original set, then we can stop detect and insert points to \tilde{Q} after we detected for each point in the coreset \tilde{P} a matching (nearest neighbour) point in \tilde{Q} that is below the required threshold.

Memory. In tracking algorithms that are based on computer vision, such as PTAM [17], we need to track hundreds of visual features. Tracking only small subset of them will save us not only time, but also memory.

Streaming and distributed Computation. Our coreset construction is applied in a streaming fashion (see Algorithm 1) using one pass over the points and memory that is independent of n . In particular, the compression (reduction) of each subset of the input markers can be computed independently and in parallel on different computers or GPUs, and then merged and maybe re-compressed again. This technique is common in coresets constructions and thus we refer the reader to e.g. [10] for details.

5 Real Time Tracking System

While our coresets are small and optimal, they come with a price: unlike random sampling which takes sub-linear time to compute (without going over the markers), computing our coreset takes the same time as solving the pose estimation problem on the same frame. To this end, we use their Property (iii) in Section 3 to suggest a real-time tracking and localization system that is based on two following threads that run in parallel.

The first thread, which we run at 1-3 FPS, gets a snapshot (frame) of markers and computes the coreset for this frame. This includes marker identification, matching problem, and then computing the actual coreset for the original set of markers P and the observed set Q .

The second thread, which calculates the object’s pose, runs every frame. In our low-cost tracking system (see 8) it handles 30 FPS. This is by using the last computed coreset on the new frames, until the first thread computes a new coreset for a later frame. The assumption of this model is that, for frames that are close to each other in time, the translation and rotation of the observed set of markers will be similar to the translation and rotation of the set Q in the previous frame, up to a small error. Property (iii) guarantees that the coreset for the first frame will still be the same for the new frame.

Our experimental results in Section 8 demonstrate the tracking system using a toy quadcopter with visual and IR features.

Table 1 shows the time complexity comparison between solving the problem on our coreset, and solving it on the entire data for different steps of the localization problem, as is done e.g. in ICP. The first row of the table represents step (1), where the matching has already been computed, and what is left to compute is the optimal rotation between the two sets of points. Test (B) in section 8, demonstrates the practical improvement of the coreset in this case.

The second row represents step (2) of the localization problem, where the matching needs to be computed given the rotation. In this case, a perfect matching between a set of

size k to a set of size m , can be achieved, according to [22], in $O(\sqrt{m+k}mk \log(m+k))$ time. Without using a coresets, the size of both sets is n . When using a coresets, the size of P is reduced to rd , although the size of Q remains n .

The last row represents a case where we need to compute the matching between two sets of points and the correct alignment is not given. In this case there are $n!$ possible permutations of the original set, each with its own optimal rotation. Using the coresets, the number of permutations reduces to roughly $(rd)!$ since it suffices to match correctly only the coresets points. Test (A) in section 8 demonstrates an actual test in which we have a noisy matching.

Table 1: Time comparison. All the numbers written in the table are in O notation and represent time complexity.

	Without using coresets $ P = n, Q = n$	Using coresets $ P = rd$
With matching, without rotation	nd^2	$ Q = rd$ d^3r
Without matching, with rotation	$n^{2.5} \log(n)$	$ Q = n$ $n^{1.5}dr \log(n)$
Noisy matching	$nd^2 \cdot (n!)$	$ Q = rd$ $(dr)!$

6 Pose-Estimation

Suppose that we have a matrix $P \in \mathbb{R}^{n \times d}$ whose rows $\{p_1, \dots, p_n\}$ represent a set of n markers on a rigid body. Let $Q \in \mathbb{R}^{n \times d}$ be a corresponding matrix whose rows $\{q_1, \dots, q_n\}$ represent the location of the markers, after rotating the body in \mathbb{R}^d with some additive random noise. Recall that a rotation matrix $R \in \mathbb{R}^{d \times d}$ is an orthogonal matrix ($R^T R = I$) whose determinant is 1. We can now define the following pose estimation problem.

Definition 1 (Wahba's Problem [23]). Find a rotation matrix R^* between the two paired sets of points P and Q in \mathbb{R}^d which minimizes the following root mean squared (RMS) deviation between them,

$$\text{cost}(P, Q, R) := \sum_{i=1}^n \|p_i - Rq_i\|^2,$$

over every rotation matrix R . We denote this minimum by

$$\text{OPT}(P, Q) := \min_R \text{cost}(P, Q, R) = \text{cost}(P, Q, R^*).$$

The Kabsch algorithm [15] suggests a simple solution for Wahba's problem. Let UDV^T be a Singular Value Decomposition (SVD) of the matrix $P^T Q$. That is, $UDV^T =$

$P^T Q, U^T U = V^T V = I$, and $D \in \mathbb{R}^{d \times d}$ is a diagonal matrix whose entries are non-increasing. In addition, assume that $\det(U) \det(V) = 1$, otherwise invert the signs of one of the columns of V . Note that D is unique but there might be more than one such factorization.

Theorem 2 (Kabsch algorithm [15]). *The matrix $R^* = VU^T$ minimizes $\text{cost}(P, Q, R)$ over every rotation matrix R , i.e., $\text{OPT}(P, Q) = \text{cost}(P, Q, R^*)$.*

7 Coreset for Pose Estimation

A *distribution* is a vector $w = (w_1, \dots, w_n)$ in \mathbb{R}^n whose entries are non-negative and sum to one, i.e., a vector in the set

$$S^n := \left\{ (w_1, \dots, w_n) \in [0, 1]^n \mid \sum_{i=1}^n w_i = 1 \right\}.$$

The *sparsity* or *support* of a distribution w is the number of non-zero entries in w .

Our goal is to compute a sparse distribution $w = (w_1, \dots, w_n)$ which defines a small weighted subset $\tilde{P} = \{\sqrt{w_i} p_i \mid w_i > 0, 1 \leq i \leq n\}$ of markers in P , and its corresponding markers $\tilde{Q} = \{\sqrt{w_i} q_i \mid w_i > 0, 1 \leq i \leq n\}$ in Q , such that solving the pose estimation problem on the pair of small sets (\tilde{P}, \tilde{Q}) would yield an *optimal* solution to the original pair (P, Q) . More generally, the coreset should satisfy the properties in Section 2 that are formalized in the following definition.

Definition 3 (Coreset for pose estimation). *Let $w \in S^n$ be a distribution that defines the matrices \tilde{P} and \tilde{Q} above. Then w is a pose coreset for the pair (P, Q) if for every pair of rotation matrices $A, B \in \mathbb{R}^{d \times d}$ and every pair of vectors $\mu, \nu \in \mathbb{R}^d$ the following holds: A rotation matrix \tilde{R} that minimizes $\text{cost}(\tilde{P}A - \mu, \tilde{Q}B - \nu, R)$ over every rotation matrix R , is also optimal for $(PA - \mu, QB - \nu)$, i.e.,*

$$\text{OPT}(PA + \mu, QB + \nu) = \text{cost}(PA + \mu, QB + \nu, \tilde{R}).$$

This implies that we can use the same coreset even if the set Q is translated or rotated over time. A coreset is efficient if it is also small (i.e. the distribution vector w is sparse).

We prove that such a coreset of constant sparsity (independent of n) always exists. Moreover, we provide an algorithm that given any pair P and Q of n points, returns such a coreset w in $O(nd^2)$ time. To this end we need to understand better the set of possible optimal solutions to the pose estimation problem, using the following lemma, and to introduce the Caratheodory theorem later.

The solution R^* in Theorem 2 is in general not unique and depends on the specific chosen SVD. In particular, suppose that R^* is an optimal solution for our coreset pair (\tilde{P}, \tilde{Q}) and also optimal for the pair (P, Q) as desired. Still, using the Kabsch algorithm with a different SVD for $\tilde{P}^T \tilde{Q}$ might yield a different matrix R' which is optimal for the coreset pair (\tilde{P}, \tilde{Q}) , but not for the original pair (P, Q) . We thus need to prove the following lemma which further characterizes the set of solutions.

Algorithm 1: POSE-CORESET(P, Q)

Input: A pair of matrices $P, Q \in \mathbb{R}^{n \times d}$.
Output: A sparse pose-coreset $w = (w_1, \dots, w_n)$ for (P, Q) ; see Definition 3.

Set $d' \leftarrow r(d-1)$ where r is the rank of P
Set $UDV^T \leftarrow$ an SVD of $P^T Q$
for each $i \in \{1, \dots, n\}$ **do**
 Set p_i and q_i to be the i th row of P and Q , respectively.
 Set $m_i \in \mathbb{R}^{d'}$ as the entries of $U^T p_i^T q_i V$, excluding its diagonal and last $d-r$ rows.
 Set $w_i \leftarrow 1$
 if $i \leq d' + 1$ **then**
 | Set $\mathcal{I}_i \leftarrow i$; Set $a_i \leftarrow m_i$
 else
 $\mathcal{I}_{d'+2} \leftarrow i$

 Set $b_j \leftarrow a_j \quad \forall j \in \{1, \dots, d' + 1\}$
 Set $b_{d'+2} \leftarrow m_i$
 $(t_1, \dots, t_{d'+2}) \leftarrow \text{SUM-CORESET}(b_1, \dots, b_{d'+2})$; see Algorithm 2
 Set $a_j \leftarrow t_j b_j, \quad \forall j \in \{1, \dots, d' + 2\}$
 Set $w_{\mathcal{I}_j} \leftarrow t_j \cdot w_{\mathcal{I}_j}, \quad \forall j \in \{1, \dots, d' + 2\}$
 Set $s \in \{1, \dots, d' + 2\}$ such that $t_s = 0$
 /* s exists by Theorem 5 */
 Set $\mathcal{I}_j \leftarrow \mathcal{I}_{j+1} \quad \forall j \in \{s+1, \dots, d' + 2\}$
 Set $a_j \leftarrow a_{j+1} \quad \forall j \in \{s+1, \dots, d' + 2\}$

 return (w_1, \dots, w_n)

Recall that UDV^T is the SVD of $P^T Q$, and let r denote the rank of $P^T Q$, i.e., number of non-zero entries in the matrix of D . Let $D_r \in \mathbb{R}^{d \times d}$ denote the diagonal matrix whose diagonal is 1 in its first r entries, and 0 otherwise.

Lemma 4. *Let $R = GF^T$ be a rotation matrix, such that F and G are orthogonal matrices, and $GD_r F^T = VD_r U^T$. then R is an optimal rotation, i.e.,*

$$\text{OPT}(P, Q) = \text{cost}(P, Q, R).$$

Moreover, the matrix $VD_r U^T$ is unique and independent of the chosen Singular Value Decomposition UDV^T of $P^T Q$.

Proof. It is easy to prove that R is optimal, if

$$\text{Tr}(RP^T Q) = \text{Tr}(D); \quad (2)$$

see [16] for details. Indeed, the trace of the matrix $RP^T Q$ is

$$\begin{aligned} \text{Tr}(RP^T Q) &= \text{Tr}(RUDV^T) = \text{Tr}(GF^T(UDV^T)) \\ &= \text{Tr}(GD_r F^T \cdot UDV^T) \end{aligned} \quad (3)$$

$$+ \text{Tr}(G(I - D_r)F^T \cdot UDV^T). \quad (4)$$

Algorithm 2: SUM-CORESET($b_1, \dots, b_{d'+2}$)

Input: Vectors $b_1, \dots, b_{d'+2}$ in $\mathbb{R}^{d'}$.

Output: Distribution $(t_1, \dots, t_{d'+2})$ of sparsity $d' + 1$ such that $\frac{1}{d'+2} \sum_i b_i = \sum_{i=1}^{d'+2} t_i b_i$.

for each $i \in \{1, \dots, d' + 2\}$ **do**

Set $u_i \leftarrow b_i - \frac{1}{d'+2} \sum_{j=1}^{d'+2} b_j$
Set $v_i \leftarrow u_i / \|u_i\|$
Set $v_i \leftarrow v_i - v_1$

Set $z_i \leftarrow \|u_i\| / \sum_{j=1}^{d'+2} \|u_j\|$, $\forall i \in \{1, \dots, d' + 2\}$

$HES^T \leftarrow \text{SVD of } [v_2 \dots v_{d'+2}]^T$

$(h_2, \dots, h_{d'+2})^T \leftarrow \text{the rightmost column of } H$

$h_1 \leftarrow -\sum_{i=2}^{d'+2} h_i$

$\alpha \leftarrow \min \left\{ \frac{z_i}{h_i} \mid 1 \leq i \leq d' + 2, h_i > 0 \right\}$

for each $i \in \{1, \dots, d' + 2\}$ **do**

Set $t'_i \leftarrow z_i - \alpha h_i$
Set

$$t_i \leftarrow \frac{t'_i}{\|u_i\|} \cdot \frac{1}{\sum_{k=1}^{d'+2} \frac{t'_k}{\|u_k\|}}$$

return $(t_1, \dots, t_{d'+2})$

Term (3) equals

$$\text{Tr}(GD_r F^T \cdot UDV^T) = \text{Tr}(VD_r U^T \cdot UDV^T) = \text{Tr}(VDV^T) = \text{Tr}(DV^T V) = \text{Tr}(D), \quad (5)$$

where the last equality holds since the trace is invariant under cyclic permutations.

Term (4) equals

$$\begin{aligned} \text{Tr}(G(I - D_r)F^T \cdot UDV^T) &= \text{Tr}(G(I - D_r)F^T \cdot (D_r U^T)^T DV^T) \\ &= \text{Tr}(G(I - D_r)F^T \cdot (V^T GD_r F^T)^T DV^T) = \text{Tr}(G(I - D_r)F^T \cdot FD_r^T G^T V \cdot DV^T) \\ &= \text{Tr}(G \cdot (I - D_r)D_r \cdot G^T V \cdot DV^T) = 0, \end{aligned}$$

where the last equality follows since the matrix $(I - D_r)D_r$ has only zero entries. Plugging the last equality and (5) in (3) yields $\text{Tr}(RP^T Q) = \text{Tr}(D)$. Using this and (2) we have that R is optimal.

For the uniqueness of the matrix $VD_r U^T$, observe that for $N = P^T Q = UDV^T$ we have

$$(N^T N)^{1/2} (N)^+ = (VDV^T)(VD^+ U^T) = VD_r U^T. \quad (6)$$

Here, a squared root $X^{1/2}$ for a matrix X is a matrix such that $(X^{1/2})^2 = X$, and X^+ denote the pseudo inverse of X . Let $FE G^T$ be an SVD of N . Similarly to (6), $(N^T N)^{1/2} (N)^+ = GD_r F^T$.

Since $N^T N = VD^2 V^T$ is a positive-semidefinite matrix, it has a unique square root. Since the pseudo inverse of a matrix is also unique, we conclude that $(N^T N)^{1/2} (N)^+$ is unique, and thus $VD_r U^T = GD_r F^T$.

The next ingredient of our coreset construction is the Caratheodory theorem [6, 9]. Recall that the convex hull of a set of points P in $\mathbb{R}^{d'}$ is the minimal convex set of $\mathbb{R}^{d'}$ that contains P .

Theorem 5 (Caratheodory’s Theorem [6, 9]). *If the point $x \in \mathbb{R}^{d'}$ lies in the convex hull of a set $P \subseteq \mathbb{R}^{d'}$, then there is a subset P' of P consisting of $d' + 1$ or fewer points such that x lies in the convex hull of P' .*

Overview of Algorithm 1. To obtain an object’s pose, we need to apply the Kabsch algorithm on the matrix $P^T Q = \sum_i p_i^T q_i$, calculating this sum takes $O(nd^2)$ time. We want to replace the summation of those n vectors by a sum of $r(d - 1) + 1$ weighted vectors $\sum_i w_i p_i^T q_i$, where w is a *pose-coreset* of sparsity at most $r(d - 1) + 1$ computed in Algorithm 1. This is done by choosing w such that

$$E = U^T \left(\sum_i w_i p_i q_i^T \right) V = \sum_i w_i (U^T p_i q_i V) \quad (7)$$

is a diagonal matrix. In this case, the rotation matrix of the pairs $(\sqrt{w_i} p_i, \sqrt{w_i} q_i)_{i=1}^n$ and (P, Q) will be the same by Theorem 2. By letting $m_i = (U^T p_i q_i V)$ we need to approximate the sum $\sum_{i=1}^n m_i$ by a weighted subset of the same sum.

Algorithm 1 reads the first $d' + 2$ points $m_1, \dots, m_{d'+2}$ and removes one of those points using Algorithm 2 that assign new weights $t_1, \dots, t_{d'+2}$ to the points such that the new weighted sum is the same, but one of the weights is zero (t_s for some s between 1 and $d' + 2$). Then the next point is inserted and another one is removed, untill all the n points m_1, \dots, m_n were inserted. The output is the final weighted set of $d' + 1$ points.

Overview of Algorithm 2. This algorithm gets a weighted set of $d' + 2$ independent vectors $b_1, \dots, b_{d'+2}$ and returns a weighted set of $d' + 1$ vectors whose weighted sum is the same. The idea is based on the proof of Theorem 5. The mean of the input vectors equals to a convex combination $\sum_i z_i b_i$, where $\sum_i z_i = 1$. We compute (possibly negative) coefficients $h_1, \dots, h_{d'+2}$ which have the same mean, $\sum_i \mu_i x_i = \sum_i z_i b_i$ and whose sum is zero. Note that adding the weights $h = (h_1, \dots, h_{d'+2})$ to the original weight vector z yields a set of weights whose sum is one and whose weighted mean is the same. This also holds for any scaling of h by $\alpha > 0$. There are $d' + 2$ values of α such that one of the weights in the sum $z + \alpha h$ will turn into zero. Taking the smallest of those values for α will make sure that the other entries are still positive; see Fig 1. The result is a desired non-negative weighted set of vectors whose weighted mean is the same.

We are now ready to prove the main theorem of this paper.

Theorem 6. *Let $P, Q \in \mathbb{R}^{n \times d}$ be a pair of matrices. Let r denote the rank of the matrix P . Then a call to the procedure $\text{POSE-CORESET}(P, Q)$ returns a pose-coreset $w \in S^n$ of sparsity at most $r(d - 1) + 1$ for (P, Q) in $O(nd^2)$ time; see Definition 3.*

Proof. Let $w \in S^n$ be a distribution of sparsity at most $r(d - 1) + 1$, and suppose that the matrix

$$E = U^T \left(\sum_i w_i p_i q_i^T \right) V \quad (8)$$

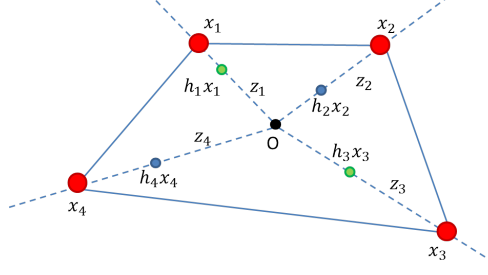


Fig. 1: Algorithm 2 computes a set of vectors (small points) whose weighted mean o (in black) is the same as the input set (large red points). The sum of those positive weights (in blue) and negative weights (green) is zero. This set is scaled by the minimal value α until one of its green points intersects a red point. The total weight of this point is 0, the other total weights are positive, and the new weighted mean is still o as desired.

is diagonal and consists of at most r non-zero entries. Here p_i and q_i are columns vectors which represent the i th row of P and Q respectively. Let $\{\sqrt{w_i}p_i \mid w_i > 0\}$ and $\{\sqrt{w_i}q_i \mid w_i > 0\}$ be the rows of \tilde{P} and \tilde{Q} respectively. Let $FE G^T$ be an SVD of $A^T \tilde{P}^T \tilde{Q} B$ such that $\det(F) \det(G) = 1$, and let $\tilde{R} = GF^T$ be an optimal rotation of this pair; see Theorem 2. We need to prove that

$$\text{OPT}(PA + \mu, QB + \nu) = \text{cost}(PA + \mu, QB + \nu, \tilde{R}).$$

We assume without loss of generality that $\mu = \nu = 0$, since translating the pair of matrices does not change the optimal rotation between them [16].

By (8), UEV^T is an SVD of $\tilde{P}^T \tilde{Q}$, and thus $A^T U E V^T B$ is an SVD of $A^T \tilde{P}^T \tilde{Q} B$. Replacing P and Q with $\tilde{P}A$ and $\tilde{Q}B$ respectively in Lemma 4 we have that $GD_r F^T = B^T V D_r U^T A$. Note that since UDV^T is an SVD of $P^T Q$, we have that $A^T U D V^T B$ is an SVD of $A^T P^T Q B$. Using this in Lemma 4 with PA and QB instead of P and Q respectively yields that $\tilde{R} = GF^T$ is an optimal rotation for the pair (PA, QB) as desired, i.e.,

$$\text{OPT}(PA, QB) = \text{cost}(PA, QB, \tilde{R}).$$

It is left to compute a sparse coreset w as defined in (8). Since P is of rank r , its rows are contained in an r -dimensional subspace. Without loss of generality, we thus assume that the last $d - r$ entries of every row p_i in P are zeros, otherwise we rotate the coordinate system. For every $i \in \{1, \dots, n\}$ let $m_i \in \mathbb{R}^{r(d-1)}$ be a vector that consists of the entries of the matrix $U^T p_i q_i^T V$, excluding its last $d - r$ rows and diagonal.

Let $u_i = m_i - \sum_{j=1}^n m_j / n$ be the translation of each m_i / n , and $v_i = u_i / \|u_i\|$. Each vector v_i is a unit vector on the unit ball, and thus the convex combination

$$(0, \dots, 0) = \sum_i \frac{\|u_i\|}{\sum_j \|u_j\|} \cdot v_i$$

lies in the convex hull of the set $\{v_1, \dots, v_n\}$. Applying Theorem 5 with $x = (0, \dots, 0)$ yields that there is a distribution $w' \in S^n$ of sparsity $r(d - 1) + 1$ such that $0 = \sum_{i=1}^n w'_i v_i$. By defining

$$w_i = \frac{1}{\sum_{k=1}^n \frac{w'_k}{\|u_k\|}} \cdot \frac{w'_i}{\|u_i\|},$$

we obtain $w \in S^n$ such that

$$\begin{aligned} \sum_i w_i m_i &= \frac{1}{\sum_{k=1}^n \frac{w'_k}{\|u_k\|}} \cdot \sum_{i=1}^n \frac{w'_i}{\|u_i\|} m_i = \frac{1}{\sum_{k=1}^n \frac{w'_k}{\|u_k\|}} \cdot \sum_{i=1}^n \frac{w'_i}{\|u_i\|} (u_i + \sum_{j=1}^n m_j / n) \\ &= \frac{1}{\sum_{k=1}^n \frac{w'_k}{\|u_k\|}} \cdot \sum_{i=1}^n w'_i v_i + \sum_{j=1}^n m_j = \sum_{j=1}^n m_j. \end{aligned}$$

Hence, the non-diagonal entries of $\sum_i w_i U^T p_i q_i^T V$ are the same as the non-diagonal entries of $\sum_{i=1}^n U^T p_i q_i^T V = D$, which are all zeros. That is, there is a diagonal matrix E such that

$$\sum_i w_i U^T p_i q_i^T V = E,$$

which satisfied (8) as desired.

$O(nd^{O(1)})$ **time implementation.** To compute w we first need to compute SVD for $P^T Q$, to get the matrices U and V , which takes $O(nd^2)$ time. Computing the set of vectors $\{v_i\}_{i=1}^n$ takes additional $O(nd)$ time. Next we need to apply the Caratheodory Theorem on this set. The proof in [6, 9] is constructive and removes a vector from the set until only the small set of $r(d-1) + 1$ vectors is left. Since each iteration takes $O(nd^2)$ time, this will take $O(n^2 d^2)$ time. To obtain a running time of $O(nd^2)$, we apply the algorithm only on $r(d-1) + 2$ vectors and run a single iteration to get $r(d-1) + 1$ vectors again; see Algorithm 1. Then, we continue to add the next vector and re-compress, until we are left with a distribution over $r(d-1) + 1$ vectors that approximates all the set, as implemented in Algorithm 2.

An SVD is not unique. The optimal rotation R^* in Theorem 2 depends on the specific chosen SVD, and since the SVD is in general not unique, we proved that the optimal rotation matrix obtained using our coresets is also optimal for the original sets even if a different SVD was chosen, see Lemma 4.

8 Experimental Results

In this section, we conducted experiments on data sets collected from real hovering tests using a sensorless, light-weight toy quadcopter, and compared the rotation error in degrees by computing the orientation of the quadcopter using our coresets as compared to uniform random sampling of markers on the quadcopter. The ground truth in the first test is an OptiTrack system, whilst in the second test the ground truth is the orientation when computed from all the markers on the quadcopter. Both the coresets and the random subset are of equal size. All our experiments show that the coresets are consistently and significantly better than a corresponding uniform sample of the same size. The error in all the experiments is measured by differences in degrees from the estimated angles (Roll, Pitch, Yaw) to the ground truth.

In both tests, the coresets were computed every x frames, the random points were also sampled every x frames, where x is called the calculation cycle. The chosen weighted points were used for the next x frames, and then a new coresets of size $5 = r(d-1) + 1$ was computed by Algorithm 1, where $d = 3$ and $r = 2$ as the LEDs on the quadcopter

are roughly placed in a planar configuration.

See attached video for demonstrations and results: <https://vimeo.com/154337055>.

8.1 Few Infra-Red Markers

In our experiments we develop a low-cost home-made tracking system ($< \$100$) that has the ability to replace commercial systems in the mission of tracking and controlling toy micro quadcopters, without any on-board sensors, assuming no high-speed maneuvers are performed (that require a professional and expensive hardware such as Vicon or OptiTrack). Our suggested tracking system requires only 2 components: (A) a mini-computer such as Raspberry Pi ($< \$30$ [2]), or a laptop/desktop , and (B) a pair of standard web-cams ($< \$5$ [3]).

The workspace covered by our system, when using two webcams, is roughly the same as the OptiTrack's workspace using two OptiTrack Flex3 cameras. Regarding the cameras' specifications, our cameras have a FOV of 56° while the OptiTrack cameras have a FOV of 58° . However, OptiTrack system obviously outperforms our low-cost system in applications that require fast maneuvers.

In order for the system to control a toy micro quadcopter, it needs to compute the $6DoF$ of the quadcopter fast enough. The sensorless quadcopter requires a stream of at least 30 control commands per second in order to hover and not crash. Using a SIFT feature detector in such hardware conditions was not practical. For faster detection in each frame, we placed 10 Infrared LEDs on the quadcopter and modified the web-cams' lenses to let only infrared spectrum rays pass. We could not place more than 10 LEDs on such a micro quadcopter since solving the matching problem on $n > 10$ markers is impractical, and due to over-weight problem.

We computed the 3D location of each LED using triangulation. Afterwards, we computed a coresets for those 3D locations, and sampled a random subset of the same size. The ground truth in this test was obtained from the OptiTrack system. The control of the quadcopter based on its positioning was done using a simple PID controller.

For different calculation cycles, we computed the average error through out the whole test, which consisted of roughly 4500 frames. The results are shown in 3b.



(a) **(left)** A toy micro-quadcopter and the 10 IR markers as (b) A toy micro-Quadcopter with a planar pattern captured by the web-camera with the IR filter **(right)**. (printed text and other features) placed on top.

Fig. 2: Toy micro-quadcopters used in our tests.

8.2 Many Visual Features

In this test, we placed a simple planar pattern on a quadcopter, as shown in Fig. 2b. An OptiTrack motion capture system was used to perform an autonomous hover with the quadcopter, whilst two other 2D grayscale cameras mounted above the quadcopter collected and tracked features from the pattern using SIFT feature detector; See submitted video. The matching between the SIFT features in both images is considered a noisy matching. This is discussed in Section 4. Given 2D points from two calibrated cameras, we were able to compute the 3D location of each detected feature using triangulation. A coresets, as described in Section 7, was computed from those 3D locations, along side a uniform random sample of the same size. The quadcopter’s orientation was then estimated by computing the optimal rotation matrix, using Kabsch algorithm, on both the coresets and the random sampled points. The ground truth in this test was obtained using Kabsch algorithm on all the points in the current frame.

For different calculation cycles, we computed the average error through out the whole test, which consisted of roughly 3000 frames, as shown in Fig. 3a. The number of detected SIFT features in each frame ranged from around 60 to 100 features, though most of the features did not last for more than 15 consequent frames, therefore we tested the coresets with calculation cycles in the range 1 to 15.

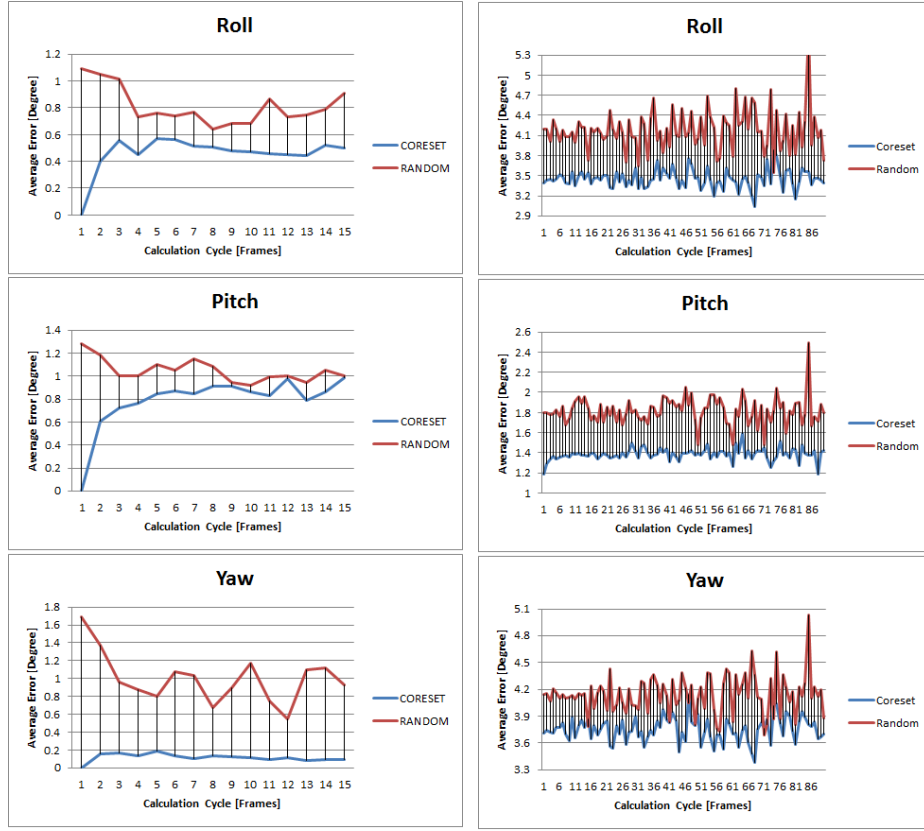
We can see that the average errors in this test were smaller than the average errors in the previous test, this is due to the low-cost hardware in the previous test, e.g. \$5 webcams as compared to OptiTrack’s \$1000 cameras, and due to the difference between the ground truth measurements in the two tests.

8.3 Time Complexity

In this section, time comparison tests have been conducted to demonstrate this paper’s main practical results. The data in this test was randomly and uniformly sampled using MATLAB. The tests were conducted on a laptop with an Intel Core i7-4710HQ CPU @ 2.50GHz processor running Ubuntu 14.04 OS.

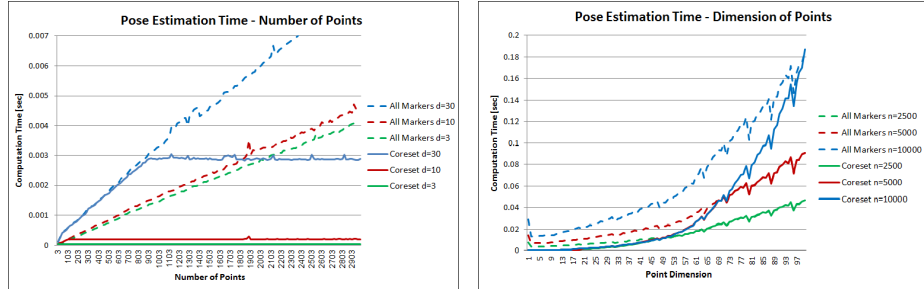
The first test compares the time needed to calculate the pose estimation given a coresets VS. using the full set of points. For this test we assume the matching between the points is given. The test has two different cases: a) Using an increasing number of points while maintaining a constant dimension, b) Using a constant number of points in many different dimensions. The results are shown in figures 3c and 3d respectively. This test is described by the first row of Table 1. Figure 3c indeed shows that when the coresets size is bigger than the number of points n , the computation time is roughly identical, and as n reaches beyond $d^2 = O(d^2)$ we see that the computation time using the full set of points continues to grow linearly with n ($O(nd^2)$), while the computation time using the coresets ceases to increase since it is independent of n ($d^3r = O(d^4)$). Figure 3d shows that the coresets indeed yields smaller computation times compared to the full set of points when the dimension $d < \sqrt{n}$, and both yield roughly the same computation time as d reaches \sqrt{n} and beyond.

The second test compares the time needed to calculate the full pose estimation problem, including the perfect matching step. For this test we assume the matching is noisy, as described in section 4 This test is described by the third row of Table 1.



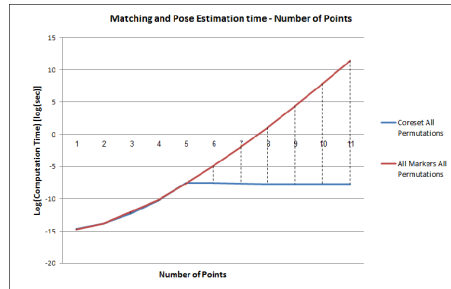
(a) For every calculation cycle (X-axis), we compare between the core-set average error and the uniform random sampling average error. The Y-axis shows the whole test (3000 frames) average error for each calculation cycle.

(b) For every calculation cycle (X-axis), we compare between the core-set average error and the uniform random sampling average error. The Y-axis shows the whole test average error for each calculation cycle.



(c) Time comparison between calculating the orientation of n points of dimension d given a previously calculated coreset versus using all n markers. The X-axis represents the number of points while the Y-axis represents the Time needed to obtain the orientation. The test was repeated for three different dimension values.

(d) Time comparison between calculating the orientation of n points of dimension d given a previously calculated coreset versus using all n markers. The X-axis represents the points dimension while the Y-axis represents the Time needed to obtain the orientation. The test was repeated for three different values of n .



(e) Time comparison between calculating the perfect matching and the orientation of n points of dimension $d = 3$ given a previously calculated coreset versus using all n markers. The X-axis represents the number of points while the Y-axis represents the $\log(\text{Time})$ needed.

Fig. 3: Experimental Results Graphs.

Bibliography

- [1] Iotracker, internet of things based tracking system. http://on_publication, 2015. To be published with a final version of this paper.
- [2] Amazon. Raspberry Pi Model A+ (256MB). <http://www.amazon.com/Raspberry-Pi-Model-A-256MB/dp/B00PEX05TO/>, 2015. [Online; accessed 15-September-2015].
- [3] Amazon. VAlinks USB 2.0 Webcam Web Cam Camera. <http://www.amazon.com/VAlinks%C2%AE-Microphone-Adjustable-Computer-Supports/dp/B0135KJWTQ/>, 2015. [Online; accessed 15-September-2015].
- [4] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Computer vision—ECCV 2006*, pages 404–417. Springer, 2006.
- [5] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992.
- [6] Constantin Carathéodory. Über den variabilitätsbereich der fourierschen konstanten von positiven harmonischen funktionen. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 32(1):193–217, 1911.
- [7] L Paul Chew, Michael T Goodrich, Daniel P Huttenlocher, Klara Kedem, Jon M Kleinberg, and Dina Kravets. Geometric pattern matching under euclidean motion. *Computational Geometry*, 7(1):113–124, 1997.
- [8] Sunglok Choi, Taemin Kim, and Wonpil Yu. Performance evaluation of ransac family. *Journal of Computer Vision*, 24(3):271–300, 1997.
- [9] Jürgen Eckhoff et al. Helly, radon, and carathéodory type theorems. *Handbook of convex geometry*, pages 389–448, 1993.
- [10] Dan Feldman, Matthew Faulkner, and Andreas Krause. Scalable training of mixture models via coresets. In J. Shawe-Taylor, R.S. Zemel, P. Bartlett, F.C.N. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 2142–2150. 2011.
- [11] Jerome H Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):209–226, 1977.
- [12] Gene H Golub and Christian Reinsch. Singular value decomposition and least squares solutions. *Numerische mathematik*, 14(5):403–420, 1970.
- [13] JF Guerrero-Castellanos, H Madrigal-Sastre, S Durand, N Marchand, WF Guerrero-Sánchez, and BB Salmerón. Design and implementation of an attitude and heading reference system (ahrs). In *Electrical Engineering Computing Science and Automatic Control (CCE), 2011 8th International Conference on*, pages 1–5. IEEE, 2011.
- [14] Richard I Hartley and Peter Sturm. Triangulation. *Computer vision and image understanding*, 68(2):146–157, 1997.
- [15] Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(5):922–923, 1976.
- [16] Hans Martin Kjer and Jakob Wilm. *Evaluation of surface registration algorithms for PET motion correction*. PhD thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark, 2010.
- [17] Georg Klein and David Murray. Parallel tracking and mapping for small AR workspaces. In *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR’07)*, Nara, Japan, November 2007.

- [18] Jung Keun Lee and Edward J Park. A minimum-order kalman filter for ambulatory real-time human body orientation tracking. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3565–3570. IEEE, 2009.
- [19] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
- [20] Kari Pulli and Linda G Shapiro. Surface reconstruction and display from range and color data. *Graphical Models*, 62(3):165–201, 2000.
- [21] M Jordan Stanway and James C Kinsey. Rotation identification in geometric algebra: Theory and application to the navigation of underwater robots in the field. *Journal of Field Robotics*, 2015.
- [22] Carlos E. Valencia and Marcos C. Vargas. Optimum matchings in weighted bipartite graphs. *Boletn de la Sociedad Matemtica Mexicana*, 2015.
- [23] Grace Wahba. A least squares estimate of satellite attitude. *SIAM review*, 7(3):409–409, 1965.
- [24] Lu Wang and Xiaopeng Sun. Comparisons of iterative closest point algorithms. In *Ubiquitous Computing Application and Wireless Sensor*, pages 649–655. Springer, 2015.
- [25] Zhengyou Zhang. Iterative point matching for registration of free-form curves and surfaces. *International journal of computer vision*, 13(2):119–152, 1994.